COMPRENSIÓN DE PROFESORES DE LA DISTRIBUCIÓN NORMAL

Yanet Karina González Arellano, Ana María Ojeda Salazar, Juan Luis Palacios Soto
Cinvestav, Universidad Autónoma Metropolitana. (México)
ygonzalez@cinvestav.mx, amojeda@cinvestav.mx, palacios.s.j.l@gmail.com

Resumen
Dos exploraciones sobre la comprensión de la distribución normal por un grupo de 26 estudiantes de bachillerato y de 11 de nivel superior (González y Ojeda, 2017) condujeron a entrevistar, en formato semiestructurado, a los dos profesores de estos estudiantes. Se les planteó un problema análogo al presentado a sus estudiantes, pero no lograron explicar la estandarización. Sus respuestas revelaron que las deficiencias de sus estudiantes en el tema procedían de una enseñanza centrada en la operatividad. Los resultados sugieren *que el estudiante repite lo enseñado y lo acepta* porque el profesor lo dice, y que el profesor también requiere, además del dominio operativo, del funcional y del analógico del concepto implicado.

Palabras clave: distribución, normal, profesores, estandarización.

Abstract
Two investigations on the understanding of the normal distribution by a group of 26 high school students and 11 university students (González and Ojeda, 2017) led to interview, in a semi-structured format, the two teachers of these students. They were posed a problem similar to that presented to their students, but they failed to explain standardization. Their answers revealed that the deficiencies their students had in the subject came from a teaching focused on the operativeness. The outcomes suggest that the student repeats what is taught, and they accept it because it is said by the teacher. So, the teacher also requires, in addition to the operational domain, the functional and analogical domain of the concept that is involved.

Key words: distribution, normal, teachers, standardization.

Introducción

Algunas investigaciones recientes acerca de la comprensión de estudiantes de la distribución normal han señalado la importancia de su enseñanza en probabilidad y estadística, así como dificultades en la transición del análisis de datos a la inferencia estadística y la escasa capacidad de análisis y síntesis al interpretar ciertos gráficos y resúmenes estadísticos (por ejemplo, Batanero, Tauber y Sánchez; 2001, 2004). Alvarado y Batanero (2007) muestran que existe una complejidad en la comprensión de una aproximación normal a la distribución binomial y sugieren reforzar la comprensión de corrección de continuidad y dedicar más tiempo al tema.

Esta investigación, cualitativa (Borovcnik, 2014), partió de los resultados de otras dos relacionadas con la comprensión de la distribución normal por estudiantes de bachillerato y de licenciatura (González y Ojeda, 2017). Los resultados de González y Ojeda (2017) exhiben deficiencias en el desempeño de los estudiantes de bachillerato, tanto en otros conceptos matemáticos necesarios para introducirlos al tema de la normal, como en los de estocásticos propios de la distribución normal. Sorprendentemente, de los estudiantes de licenciatura se obtuvieron casi los mismos resultados. Por ello investigamos la comprensión de los dos profesores respectivos para determinar si las respuestas de sus estudiantes apuntaban a errores propios o a interpretaciones deficientes derivadas de su enseñanza.

Elementos teóricos

De las diez ideas de estocásticos como guía continua de un currículum en espiral, propuestas por Heitele (1975) con un enfoque epistemológico, seis están implicadas en la distribución normal: medida de probabilidad, espacio muestra, adición de probabilidades, equiprobabilidad y simetría, variable estocástica y muestra. Para la idea de variable estocástica, el autor señaló su importancia respecto a tres puntos: *la distribución de una variable, su esperanza y la composición de variables estocásticas para obtener otras nuevas*. En particular, la estandarización consiste en un cambio de la variable X a Z donde Z está en función de X, por lo que la probabilidad de X se calcula a través de Z.

Steinbring (1991) propone un triángulo epistemológico para explicar la constitución de un concepto matemático, como las interrelaciones entre éste, el objeto (contexto de referencia) y su signo. Así indica que objeto, signo y concepto, no pueden ser tratados de forma independiente en la deducción del significado del conocimiento matemático.

Pollatsek, Lima y Well (1981) señalan que para considerar la comprensión de un concepto matemático se tomen en cuenta los tres tipos de conocimiento: *de cálculo* (corresponde a lo operativo), *conocimiento funcional* (hace referencia a la forma en que se aplica el concepto en una situación particular) y *conocimiento analógico* (a manera de uso de metáforas, puede incluir imágenes visuales que contribuyan a evitar que el estudiante cometa errores).

Gigerenzer y Hoffrage (1995) indican cómo, en general, la estimación de probabilidades privilegia el formato de frecuencia sobre el de probabilidad, lo que puede ocasionar una desatención a la tasa base.

Hogarth (2001) sugiere fomentar la interacción entre el sistema deliberado (exige esfuerzo y atención) y el tácito (sin atención consciente o esfuerzo) en la enseñanza de un concepto. Para él, la intuición es el resultado de las experiencias a las que nos sometemos; su esencia de la intuición o de las respuestas intuitivas radica en que se alcanzan con poco esfuerzo y sin una conciencia deliberada, por lo cual recomienda educar la intuición para evitar intuiciones incorrectas y ser más eficiente en la toma de decisiones.

Método e instrumentos

Entrevistamos individualmente con un guión semiestructurado (Zazkis y Hazzan, 1999) a los dos profesores que enseñaron el tema de distribución normal a los estudiantes de los grupos investigados de bachillerato (González y Ojeda, 2017) y de licenciatura. El profesor D_b, de 36 años, había enseñado Probabilidad y Estadística a nivel bachillerato, era Licenciado en Matemáticas Aplicadas y Computación, y ya había impartido la materia dos veces. El profesor D_S, de 46 años, había impartido materias de Estadística y Probabilidad a nivel superior desde hacía tres años en distintas áreas de ciencias básicas (física, computación), ciencias sociales y humanidades, así como ciencias biológicas; era Licenciado y Maestro en Matemáticas y estaba por concluir un Doctorado en Matemáticas. A los dos se les plantearon preguntas similares a las propuestas a sus estudiantes. El objetivo fue identificar posibles rasgos de sus enseñanzas del tema, en particular qué tipos de conocimiento (Pollatsek et al., 1981) habrían favorecido en ellas; también se les preguntó acerca de su forma de enseñar el tema de la distribución normal, una descripción general de lo que enseñaban en sus cursos acerca del tema y los libros de los que lo documentaban. Además, nos interesó si el profesor *también repetía* el discurso del libro de su referencia. En las entrevistas, videograbadas y transcritas para su análisis, se les planteó este problema en formato digital, de la obra de Gutiérrez y Vladimirovna (2014, pp. 210):

Ciertos tipos de baterías para automóvil tienen un tiempo de vida normalmente distribuido con media 1200 días y desviación estándar igual a 100 días.
a) ¿Cuántas de las siguientes 3000 baterías que se venderán durarán más de1300 días?
b) ¿Por cuánto tiempo se deben garantizar las baterías si el fabricante quiere reemplazar sólo 10 por ciento de las baterías vendidas?

Se les leyó el enunciado y después se les permitió leerlo el tiempo que requirieran para describir cómo lo resolverían; en caso de hacer anotaciones disponían de pintarrón y plumones. La Tabla 1 resume la aplicación al problema de los criterios de análisis de la célula (Ojeda, 2006).

Tabla 1. *Criterios de análisis al problema planteado en entrevista.*

Criterio	**Problema**
Ideas fundamentales de estocásticos	
1) Medida de probabilidad	$P(X \in [a,b]) = \int_a^b \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} dx$. Por ejemplo $P(X > 1300) = 0.1587$.
2) Espacio muestra	$[0,\infty)$.
3) Adición de probabilidades.	Dados dos subconjuntos $A, B \subset [0,\infty)$, tal que $A \cap B = \emptyset$, $P(A \cup B) = P(A) + P(B)$. En particular se cumple esta propiedad si tomamos a $A = [0,x)$ con $x < \mu$ y $B = (1300,\infty)$, debido a que $\mu < 1300$.

4) Equiprobabilidad y simetría	La función es simétrica respecto a $\mathrm{x} = \mu = 1200$, por lo que respecto a este valor podemos encontrar eventos equiprobables.
5) Variable aleatoria X	Tiempo de vida de las baterías.
6) Muestra	3000 baterías.
Otros conceptos matemáticos	Orden en los reales, proporcionalidad, operaciones aritméticas, porcentaje, integrales, cambio de variable.
Recursos semióticos	Lengua natural escrita, simbología matemática, tablas de Z, gráfica normal.
Términos para referirse a estocásticos	Tiempo de vida, normalmente distribuido, media, desviación estándar, cuántas de 3000, más de 1300 días, cuánto tiempo, 10 por ciento.
Referente	¿Cuántas de las 3000 baterías durarán más de 1300 días?

Una dificultad que podría tener el profesor al presentarle el problema sería la interpretación del enunciado. Si X es la variable aleatoria *tiempo de vida de las baterías*, para una muestra de 3000 baterías, se requiere determinar cuántas durarán más de 1300 días.

Encontrar $P(X > 1300)$ resultaría sólo en la probabilidad de que las baterías tuvieran un tiempo de vida mayor a 1300, no en el número de baterías requerido. Pero se puede poner en correspondencia esta probabilidad con el número de baterías y establecer una biyección entre el intervalo $[0,1]$ y $[0,3000]$ mediante una función, por ejemplo $G(p) = 3000p,\ p \in [0,1]$. Al estandarizar tenemos que $P(X > 1300) = 0.1587$, por tanto, 15.87 % de las 3000 baterías durará más de 1300 horas (aproximadamente 476 baterías).

Otra forma de plantear la pregunta del inciso b) sería: ¿para qué valor $x < \mu, P(X < x) = 0.1$? Como la función de densidad es simétrica, los profesores podrían considerar el caso $P(X > x_1) = 0.1$ o, más aún, dos valores x_2 y x_3 tales que $P(x_2 < X < x_3) = 0.1$. Pero el profesor debía percatarse de que en el enunciado del problema estas dos últimas expresiones no tienen sentido. Por ejemplo, el caso $P(X > x_1) = 0.1$ garantizaría un tiempo de vida mayor a la media.

Para normalizar la variable X se consideró que el profesor no tendría dificultad en aplicar las propiedades de las desigualdades a $P(X < x) = 0.1$, como sucedió con estudiantes del profesor D_b; aunque sí podría faltar reflexión en el procedimiento de estandarización, como se reveló en las entrevistas a estudiantes de los dos profesores. Por ello, parte de las preguntas de la entrevista a los docentes correspondió a la estandarización para determinar si no habían favorecido su comprensión por falta de tiempo o de reflexión.

$$P(X < x) = P(X - 1200 < x - 1200) = P\left(\frac{X - 1200}{100} < \frac{x - 1200}{100}\right) = P(Z < z) = 0.1,$$

con Z la variable normal estándar, $Z = \frac{X-1200}{100}$, y z un valor particular de Z que corresponde a $z = \frac{x-1200}{100}$. Al continuar con la solución se previó una dificultad para recuperar de Z la variable original X, para finalmente determinar el valor de x que satisface $P(X < x) = 0.1$; también se previó una posible dificultad en la lectura de las tablas, aunque en ninguno de los dos casos de entrevista se llegó al uso de las mismas.

De las tablas para Z acumulada, con $z = -1.28$, resulta $P(Z < -1.28) \approx 0.1$; por lo que de la ecuación $-1.28 = \frac{x-1200}{100}$, se tiene $x = 1072$. Esto significa que el fabricante debe garantizar las baterías al menos por 1072 días para que, en caso de que duren menos de ese tiempo, reemplace aproximadamente sólo el 10 % de las baterías vendidas.

Resultados del análisis

Las Tablas 2 y 3 caracterizan la comprensión de la distribución normal de D_b y D_S, respectivamente.

Caso D_b. El profesor de bachillerato afirmó haber resuelto problemas de ese tipo en clase, como el del diámetro del tallo de girasoles. No tuvo dificultades para describir el procedimiento de solución del problema planteado, no así en la parte conceptual:
I_1: ¿Quién sería el espacio muestra de este ejercicio?

D_b: Este… el tiempo de vida de las baterías ¿no? O… ¿cómo? ¿Cómo le hacemos? No, no sé.

A la pregunta de a qué correspondía el eje de las ordenadas en la gráfica de la distribución normal para el tiempo de vida de las baterías, contestó que a los valores de la probabilidad.

Tabla 2. *Comprensión de D_b de la distribución normal*

Ideas fundamentales de estocásticos	Otros conceptos matemáticos	Recursos semióticos empleados
No logró proporcionar los posibles valores de la función de probabilidad. No recordó lo que es un espacio muestra, ni lo describió para el problema planteado. Se mostró dudoso de relacionar la variable aleatoria con el tiempo de vida de la batería. Reconoció no haberse percatado del proceso subyacente a la estandarización.	Reconoció no haber reflexionado en la relación entre puntos de inflexión y la desviación estándar. Y al igual que sus estudiantes, no podría dar el valor de la desviación a partir de la gráfica y de los puntos de inflexión.	Trazo de la curva normal en el plano cartesiano. Simbología matemática. Lengua natural escrita.

Fuentes de referencia. La obra de Mendenhall, Beaver y Beaver (2010), citada por D_b como obra de consulta, presenta en la página 235 un ejemplo similar al proporcionado por D_b, acerca del diámetro de un girasol como una aplicación de la distribución normal. Además, el libro sí presenta ejemplos similares al inciso a) (pág. 231) y al inciso b) (pág. 232) del problema planteado al profesor en la entrevista.

El libro de Johnson y Kuby (2012) presenta ejemplos (pág. 281), operativos, con conversiones de porcentaje a decimal. Además, el libro presenta la igualdad 10 % = 0.1000 (pág 281), de lo que resalta la insistencia de D_b en que sus estudiantes asignaran una probabilidad como un porcentaje:

I_1: ¿Es correcto dar una probabilidad como un porcentaje?
D_b: Pues yo digo que sí. Bueno, así yo he concluido… en porcentajes, sí… Pero sí les digo "lo mejor es escribir la conclusión en porcentajes".

Observación de sus estudiantes. D_b no se refirió a dificultades de sus estudiantes relativas a las ideas de estocásticos, pero sí señaló algunas concernientes a otros conceptos matemáticos requeridos, como la conversión de porcentajes a su expresión decimal y a los signos de orden. Anticipó dos posibles dificultades más si les planteara el problema del tiempo de vida de las baterías: porcentaje y notación científica. D_b reconoció que algunos estudiantes se inscribían a *Probabilidad y Estadística* por considerar las materias más fáciles que las de *Cálculo Diferencial e Integral*, es decir, por huir de las matemáticas.

Forma de enseñanza. Del análisis de la entrevista se evidencia que prevalece un conocimiento de cálculo, favorece el uso de procedimientos y señaló que por la limitación del tiempo no va más allá de esto. Señaló que hasta éste, su segundo curso que impartía, no había sensibilizado a sus estudiantes sobre los posibles valores que toma la función de probabilidad, además de haber limitado su enseñanza de la estandarización al uso de la fórmula de Z. Reconoció que en este segundo curso impartido había favorecido, a raíz de una pregunta en entrevista, hecha a sus dos estudiantes del curso anterior (González y Ojeda, 2017), contrastar el sombreado del área bajo la curva normal y el valor de la probabilidad obtenido.

Caso D_S. Aunque el profesor del nivel universitario se refirió a cómo se distribuyen los datos por ejemplo de las estaturas y pesos de personas y que planteó el caso particular de una persona para determinar la probabilidad de un evento de interés, al presentarle el inciso b) del problema de las baterías contestó que no sabía cómo hacerlo.

I_1: ¿Podrías sombrear la región? ¿Gráficamente qué es lo que pide el inciso b)?
D_S: … No recuerdo cómo. No me acuerdo cómo.

Tabla 3. Comprensión de D_S de la distribución normal

Ideas fundamentales de estocásticos	Otros conceptos matemáticos	Recursos semióticos empleados
Definió correctamente a una variable aleatoria y al espacio muestra. Proporcionó de forma correcta los valores de la función de probabilidad. Describió estandarización como: cuando divides entre la desviación	No se percató antes que X se trataba de una variable aleatoria continua, por lo que los extremos de integración eran los mismos. Indicó calificaba mal al estudiante si	Trazo de histogramas y curva normal y normal estándar. Simbología matemática. Lengua natural. Al preguntársele si podría relacionar la desviación estándar con la gráfica respondió que no

estándar. En el momento [en] que haces una división entre la desviación estándar de la variable aleatoria, en ese momento estandarizas. Se dice que estás estandarizando. Al pedirle que ampliara su respuesta, respondió que no podría.	escribía $P(X \geq 1300)$ en vez de $P(X > 1300)$.	se acordaba. La pregunta original apuntaba a si podía obtener información a partir de la gráfica.

I_2: ¿Qué es estandarizar? ¿Qué es lo que estás haciendo?
D_S: Bueno aquí podríamos ver que como le estoy restando a la variable, le estoy restando la media, estoy sacando una dispersión que hay entre la media y la variable aleatoria ¿no? Entonces ese numerito lo estoy dividiendo entre la desviación estándar de la variable. Entonces mmm ¿cómo podría yo explicar eso? Déjame ver… No se me ocurre.

Fuentes de referencia. De las obras consultadas por D_s para la distribución normal, señalamos lo siguiente: 1) la de Levin, Rubin, Balderas, del Valle y Gómez (1987) proveen de un buen número de ejemplos; no dan mayores detalles técnicos del proceso de estandarización. La definen sólo con su expresión simbólica, i.e. $Z = \frac{X-\mu}{\sigma}$ (pág. 213). Para relacionar toda distribución normal con los parámetros μ y σ argumentan que todas tienen la misma área bajo la curva de $-\lambda\sigma$ a $\lambda\sigma$ desviaciones estándar; proporcionan ejemplos para $\lambda = 1, 2$ y 3, pero no sustentan teóricamente su afirmación. 2) La de Freund, Miller y Miller (2000) presenta en la sección 6.5 la técnica de cambio de variable (Teorema 6.7, p. 220) y despliegan la herramienta necesaria para la estandarización (Teorema de cambio de variable). 3) Anderson, Sweeney y Williams (2008) utilizan el término *estandarización* de manera somera, no se refieren al cambio de variable y sólo señalan: "Para la distribución normal estándar ya se encuentran calculadas las áreas bajo la curva normal y se cuenta con tablas que dan estas áreas y que se usan para calcular las probabilidades" (p. 234).

Observación de sus estudiantes. D_S indicó que había cierto rechazo por un grupo de estudiantes hacia la Probabilidad y la Estadística por no considerarlas matemáticas "serias" [teóricas] y que esto obstaculizaba su enseñanza, sumado al poco tiempo del curso y a sus deficiencias en otros conceptos matemáticos, como dificultades con los signos de orden.

Forma de enseñanza. D_S señaló la importancia de la aplicación de problemas para que el estudiante entendiera el concepto de distribución normal. También que recurría al trazo de las gráficas normal y normal estándar, una encima de la otra, para que el estudiante visualizara de forma gráfica el problema en las variables X y Z. Aunque todo esto supondría la preferencia por un conocimiento al menos funcional, el docente subrayó que realizaba los cálculos minuciosamente para que los estudiantes los entendieran, por lo que también para él concluimos que prevalece en su enseñanza un conocimiento de cálculo.

Comentarios

Triángulo epistemológico. El referente del problema planteado, aunque análogo pero distinto del de los resueltos por *Ds* en clase, impidió que éste lo resolviera. Ninguno de los dos profesores advirtió la interrelación entre *objeto* (distribución del tiempo de vida de baterías), *signo* ($Z = (X - \mu)/\sigma$) y *concepto* ($N(\mu, \sigma^2)$) (Steinbring, 1991).

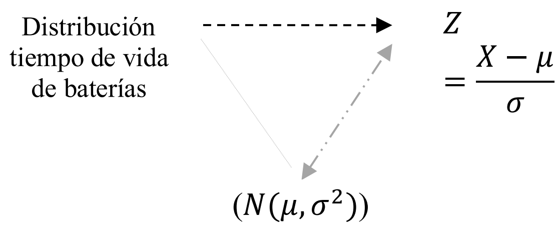

Figura 1. Aplicación a la distribución normal al triángulo epistemológico

Db no dio evidencia de comprender las ideas fundamentales de medida de probabilidad, espacio muestra, ni variable aleatoria (Heitele, 1975). El referente del problema planteado, aunque análogo a los presentados en los libros que *Ds* señaló como referencias para apoyarse en los cursos, pero distinto del de los resueltos en clase, impidió que éste lo resolviera. Ni *Db* ni *Ds* relacionaron los puntos de inflexión con los valores $\mu \pm \sigma$. *Db* no distinguió entre un enfoque probabilístico y frecuencial (Gigerenzer y Hoffrage, 1995) y reconoció haber enseñado a sus estudiantes que debían concluir una probabilidad como un porcentaje. En general, las deficiencias que los profesores mostraron fueron similares a las de sus estudiantes, lo que indica que, en acuerdo con Hogarth (2001), el estudiante sólo repite lo que su profesor le enseña, sea correcto o incorrecto, pues en general el primero no se documenta más allá de lo que se imparte en el aula. Además, se evidenció que el profesor *Db* repetía lo que presentaban sus libros de consulta, lo que reveló que no sólo el estudiante repetía el discurso del profesor, sino que el profesor repetía el discurso del libro en el que se basaba. En ambas entrevistas se reveló que sus enseñanzas habían fomentado el uso correcto de tablas y de cálculos, en vez de la reflexión, incluso de su propia reflexión. Por lo cual, en acuerdo con Pollatsek y sus colaboradores (1981), subrayamos la necesidad de favorecer en la enseñanza además los conocimientos funcional y analógico de los conceptos, no sólo el de cálculo. Más aún, es necesario romper con la consideración de que la Probabilidad y la Estadística consisten simplemente en efectuar cálculos y comenzar a poner en juego los dos tipos de pensamiento que señala Hogarth (2001).

Referencias bibliográficas

Anderson, D., Sweeney, D. y Williams, T. (2008). *Estadística para administración y economía*, 10ª. ed. México: Cengage Learning Editores.
Borovcnik, M. (2014). Empirical research on understanding probability and related concepts – A review of vital issues. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in Statistics Education. Proceedings of the Ninth International Conference on Teaching Statistics* (ICOTS9, July, 2014), Flagstaff, Arizona, USA.
Freund, J y Miller, I. (2000). *Estadística matemática con aplicaciones*, 6ª ed. México: Pearson Educación.

Gigerenzer, G. & Hoffrage, U. (1995). How to improve bayesian reasoning without instruction: Frequency formats. *Psychological Review*, 102, 684-704.

González, Y. y Ojeda, A. M. (2017). Comprensión de la distribución normal en bachillerato. *Acta Latinoamericana de Matemática Educativa 30*. México: Comité Latinoamericano de Matemática Educativa A. C.

Gutiérrez, E. y Vladimirovna, O. (2014). *Probabilidad y Estadística. Aplicaciones a la Ingeniería y Ciencias*. México: Grupo Editorial Patria.

Heitele, D. (1975). An Epistemological View on Fundamental Stochastic Ideas. *Educational Studies in Mathematics*, 6 (2), 187-205.

Hogarth, R. (2001). *Educar la intuición*. Barcelona: Paidós.

Johnson, R. y Kuby, P. (2012). *Estadística elemental*, 11ª edición. México: Cengage Learning Editores.

Levin, R. y Rubin, D. (2004). *Estadística para administración y economía*, 7ª ed. México: Pearson Educación.

Mendenhall, W., Beaver, R. y Beaver, B. (2010). *Introducción a la probabilidad y estadística*, 13ª ed. México: Cengage Learning Editores.

Pollatsek, A., Lima, S. & Well, A. (1981). Concept or Computation: Students' Understanding of the Mean. *Educational Studies in Mathematics*, 12(2), 191-204.

Steinbring, H. (1991). The Concept of Chance in Everyday Teaching: Aspects of a Social Epistemology of Mathematical Knowledge. *Educational Studies in Mathematics*, 22, 503-522.

Zazkis, R. & Hazzan, O. (1999). Interviewing in Mathematics Education Research: Choosing the Questions. *Journal of Mathematical Behavior*, 17(4), 429-439.